MOLECULAR AND CELLULAR BIOLOGY, Dec. 1991, p. 6075–6083
0270-7306/91/126075-09$02.00/0

Vol. 11, No. 12

Selection of Splice Sites in Pre-mRNAs with Short Internal Exons

ZBIGNIEW DOMINSKI AND RYSZARD KOLE*

Department of Pharmacology and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599

Received 30 May 1991/Accepted 4 September 1991

Model pre-mRNAs containing two introns and three exons, derived from the human β-globin gene, were used to study the effects of internal exon length on splice site selection. Splicing was assayed in vitro in HeLa nuclear extracts and in vivo during transient expression in transfected HeLa cells. For substrates with internal exons 87, 104, and 171 nucleotides in length, in vitro splicing proceeded via a regular splicing pathway, in which all three exons were included in the spliced product. Primary transcripts with internal exons containing 23, 29, and 33 nucleotides were spliced by an alternative pathway, in which the first exon was joined directly to the third one. The internal exon was missing from the spliced product and together with two flanking introns was included in a large lariat structure. The same patterns of splicing were retained when transcripts containing 171-, 33-, and 29-nucleotide-long internal exons were spliced in vivo. A transcript containing a 51-nucleotide-long exon was spliced in vitro via both pathways but in vivo generated only a correctly spliced product. Skipping of short internal exons was reversed both in vitro and in vivo when purines in the upstream polypyrimidine tract were replaced by pyrimidines. The changes in the polypyrimidine tract achieved by these substitutions led in vitro to complete (transcripts containing 28 pyrimidines in a row) or partial (transcripts containing 15 pyrimidines in a row) restoration of a regular splicing pathway. Splicing in vivo of these transcripts led exclusively to the spliced product containing all three exons. These results suggest that a balance between the length of the uninterrupted polypyrimidine tract and the length of the exon is an important determinant of the relative strength of the splice sites, ensuring correct splicing patterns of multiintron pre-mRNAs.

There are four sequence elements within introns of mammalian pre-mRNAs which are required for efficient splicing. These are the conserved sequence at the 5′ splice junction, the invariant AG dinucleotide at the 3′ splice junction, a weakly conserved sequence at the site of lariat formation, called the branch point sequence, and a pyrimidine-rich region of variable length between the branch point and the 3′ splice junction, referred to as the polypyrimidine [poly(Y)] tract (for reviews see references 20, 32, 43). The importance of these elements has been shown experimentally in a number of studies employing pre-mRNAs containing point mutations or deletions within existing sequences (1, 37, 41, 49, 51, 53). Although important, the four sequence elements seem insufficient to account for the high specificity of the splicing reaction. Similar sequence motifs, termed cryptic splice sites, are present in many regions throughout pre-mRNAs and yet are not recognized under normal conditions (31), whereas in transcripts bearing point mutations or deletions within natural elements, such sites can be activated (1, 37, 50, 53). Furthermore, in multi-intron primary transcripts, the mechanism of splice selection must ensure correct pairing of splice sites since recognition of the 5′ splice site from one intron and the 3′ splice site from a distant one would result in exon skipping and subsequent deletion of genetic information. The complexity of this mechanism is compounded by the existence of alternative splicing. This process leads to the generation of different protein isoforms from the same primary transcript by alternative selection of splice sites in a tissue-specific or developmentally controlled manner (reviewed in references 3 and 25). Therefore, the mechanism of splice site selection, while maintaining the high specificity which allows it to ignore cryptic splice sites and choose correct pairs of 5′ and 3′ splice sites, must remain flexible enough to distinguish between two alternative splice sites, depending on the physiological or developmental state of the cell.

The mechanism of splice site selection in transcripts allowing for *cis* competition between duplicated splice sites was investigated in several studies. It was found that the selection of 5′ splice sites depended mainly on the degree of their complementarity to the appropriate U1 small nuclear RNA sequences (24, 30, 55, 56). The distance between the duplicated 5′ splice sites was also shown to affect their selection (38). For the selection of 3′ splice sites, the sequence compositions of the branch point (and its complementarity to U2 small nuclear RNA) (39, 54, 57) and of the poly(Y) tract (11, 13, 36) were found to be important. Furthermore, in certain instances, including physiological alternative splicing, the context of 5′ and 3′ splice sites, i.e., the sequences of adjacent exons, seemed to affect splice site selection (12, 38). The role of exons in splicing was previously suggested by the results from this (14, 15) and other (33, 52) laboratories which showed that monointronic transcripts containing a short downstream exon were spliced in vitro very inefficiently.

In most of the studies cited above, duplicated splice sites at one end of a single intron competed for a single splice site at the other end. In this work, we have constructed a set of model transcripts containing two introns and three exons, with the length of the internal exon ranging from 23 to 171 nucleotides. Such transcripts contain two donor and two acceptor splice sites available for competition and approximate more closely natural multi-intron pre-mRNAs. By testing in vitro and in vivo splicing of these transcripts, we have found that the length of the internal exons plays an important role in splice site selection. Transcripts with an

* Corresponding author.

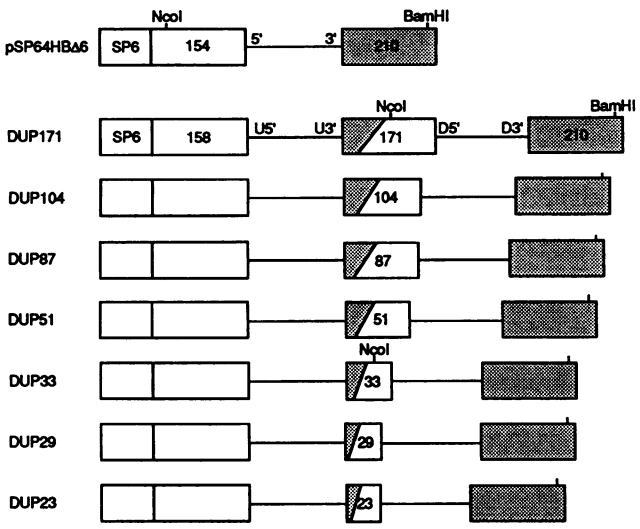

FIG. 1. Structures of DNA templates used for SP6 transcription. Boxes, exon sequences; lines, intron sequences. Numbers indicate the sizes of exons in nucleotides; both introns in all constructs are 130 nucleotides long. 5′ and 3′ splice sites in either the upstream (U) or downstream (D) introns are indicated. The first exon in clone DUP171 and in all of its derivatives is 4 nucleotides longer than the corresponding exon in pSP64HBΔ6 as a result of removal of the *Nco*I site by filling in the ends with Klenow polymerase. For transcription with SP6 RNA polymerase, all templates were linearized at the *Bam*HI site.

internal exon over 51 nucleotides in length display regular splicing pathways in which three exons are included in the spliced product. In contrast, internal exons 33 nucleotides long or shorter are almost completely ignored, leading to exon skipping as the main splicing pathway.

We have found also that skipping of the short internal exons can be reversed in vitro and in vivo when single purines in the poly(Y) tracts of the upstream introns are replaced by pyrimidines. The increased length of uninterrupted poly(Y) tracts achieved by these substitutions apparently led to improved recognition of the internal splice sites and restoration of the regular splicing pathway. These results suggest that one of the important parameters that governs splice site selection in multi-intron pre-mRNA is determined by the relationship between the length of the exons and the nature of the upstream poly(Y) tracts.

MATERIALS AND METHODS

Construction of plasmids. All plasmid manipulations were carried out according to standard methods (26). For construction of clone DUP171, the *Sph*I-*Ava*II restriction fragment of pSP64HBΔ6 (21), containing the SP6 promoter, the whole first exon, first intron, and 23 nucleotides of the second exon of the human β-globin gene was ligated with the large fragment of the same plasmid cut with restriction enzymes *Sph*I and *Hin*dIII. Prior to ligation, the *Nco*I site in the plasmid was destroyed by filling in the ends with Klenow polymerase. The structure of the resultant construct is shown in Fig. 1. Clones DUP23, DUP29, DUP51, DUP87, and DUP104 were obtained by Bal 31 nuclease digestion of plasmid DUP171 linearized at the *Nco*I site in the internal exon. For construction of clone DUP33, Bal 31 nuclease digestion was followed by blunt-end ligation in the presence of *Nco*I linker to retain this site. The structure of each construct (Fig. 1) was confirmed by sequencing. DNA clones used for in vivo studies were constructed by inserting into the *Hin*dIII site of the clones described above a *Bam*HI-*Sau*3A fragment of plasmid pHD1012 (8). This fragment carries an immediate-early promoter of human cytomegalovirus and ends 2 nucleotides downstream of the initiation of transcription site. The original, β-globin polyadenylation signal was positioned 25 nucleotides downstream from the *Bam*HI site located in the third exon of each of the constructs. The details of plasmid constructions are available on request.

Site-directed mutagenesis. To introduce single-base substitutions within the poly(Y) tract, site-directed mutagenesis was carried out as described elsewhere (22). One microgram of uracil-containing single-stranded DNA and a 10-fold molar excess of mismatched oligonucleotide were used to promote the synthesis of a new strand with Sequenase version 2.0 polymerase (U.S. Biochemical). All resulting mutants were derivatives of clone DUP33, which because of a unique *Nco*I restriction site in the internal exon allowed separation of both duplicated halves of the construct. This, in turn, facilitated introduction of point mutations into only one of the two identical introns. The following oligonucleotides were used (written in the 5′-to-3′ orientation):

```
GGTGGGAAAAAAGAGAAAAAGGGAGAGAGAGTC (DUP33Y5)
GGTGGGAAAAGAGAAGAAGAGGAAAGAGAGAGTC (DUP33Y5')
GGTGGGAAAAGAGAAGAATAGGCAGA (DUP33Y3)
AATAGACCAAGAGGAAGAGAGAGTC (DUP33Y2)
AGCCTAAGGGGGGGAAAATAG (DUP33Y1)
```

Sequences of the mutated poly(Y) tracts are shown in Fig. 7A.

In vitro transcription and splicing. ^{32}P-labeled transcripts, capped at the 5′ end, were obtained by using SP6 RNA polymerase (Promega) and DNA templates truncated at the *Bam*HI site, as previously described (18). In vitro splicing was carried out in the nuclear extract from HeLa cells (9, 21) as described elsewhere (14). The RNA from the reaction was purified by phenol extraction followed by ethanol precipitation and analyzed by electrophoresis on an 8% polyacrylamide sequencing gel.

Transfection of HeLa cells and isolation of the cellular RNA. About 10^7 HeLa cells from suspension culture grown in Joklik's modified minimal Eagle's medium, containing kanamycin and gentamicin and supplemented with 5% horse serum (Flow Laboratories, Inc.), were mixed with approximately 30 μg of DNA in a total volume of 0.5 ml. Electroporation was carried out as described previously (35). After treatment, the cells were transferred to a 100-mm Falcon 3003 tissue culture dish and grown in the same medium for 40 to 48 h. RNA was isolated by the guanidinium-cesium chloride method as described previously (26). About 50 μg of cellular RNA was obtained from cells harvested from a single dish.

Primer extension and sequencing. For primer extension analysis, a 25-mer oligonucleotide, complementary to nucleotides 76 to 100 in the second exon of the human β-globin gene, was used. Approximately 30 μg of total RNA from transfected HeLa cells and 20 to 50 ng of the primer labelled at the 5′ end by polynucleotide kinase (200,000 cpm) were annealed for 30 min at 65°C and then for 3 h at 45°C in a buffer containing 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 10 mM dithiothreitol, and 3 mM $MgCl_2$. The reaction mixture was then supplemented with deoxynucleotide triphosphates to a final concentration of 0.2 mM and with 200 U of Moloney murine leukemia virus reverse transcriptase (Be-

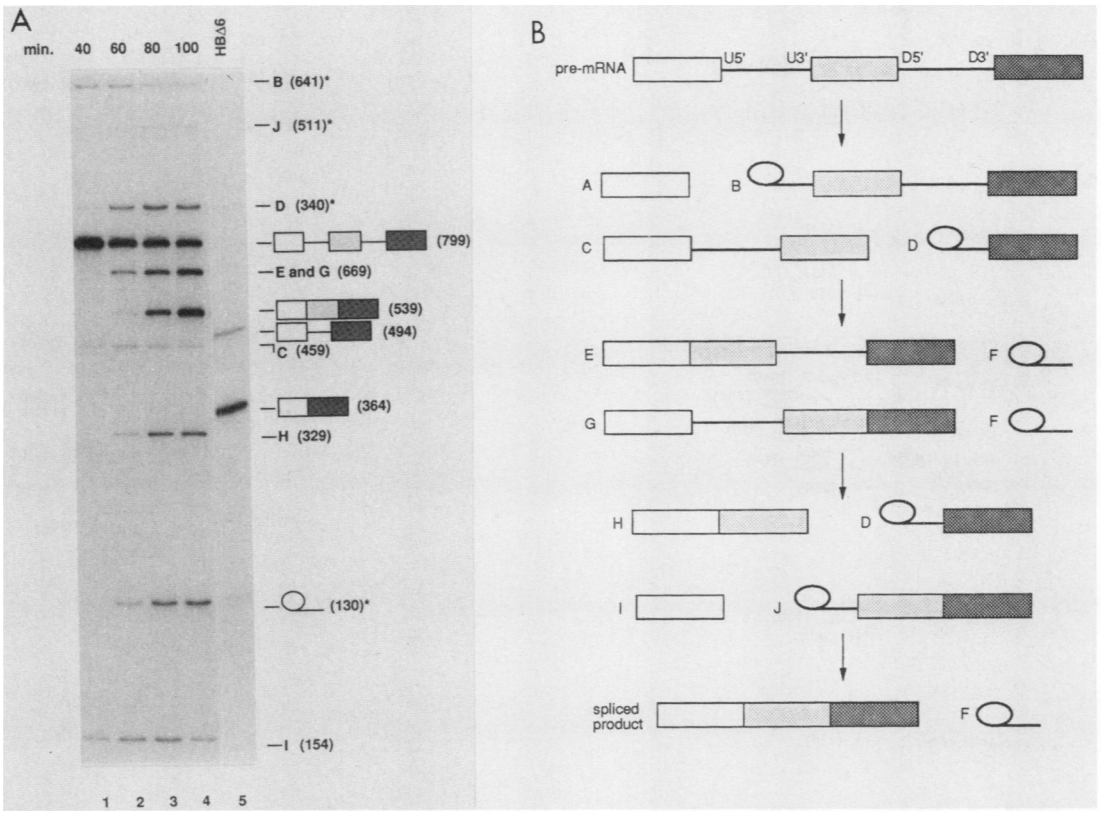

FIG. 2. (A) Time course of the in vitro splicing reaction for DUP171 pre-mRNA. A 799-nucleotide-long substrate was incubated with HeLa cell nuclear extract for the times indicated at the top and analyzed on an 8% polyacrylamide sequencing gel. As a control, in vitro splicing of HBΔ6 pre-mRNA was carried out (lane 5). Sizes (in nucleotides) and structures of all final products are depicted on the right. All intermediates generated during the reaction are designated by capital letters and appropriate size. The structures of all of the products and intermediates are illustrated in panel B. An asterisk denotes aberrant electrophoretic mobility of lariat-containing intermediates. (B) Predicted splicing pathway for a three-exon pre-mRNA proceeding by correct pairing of splice sites. Intermediates of the reaction are designated by capital letters. The same designations are used in Fig. 2A, 3, 5, and 7B.

thesda Research Laboratories) and incubated at 40°C for 60 min (total volume, 40 μl). The samples were treated with RNase A (30 μg/ml) for 10 min at 37°C, extracted twice with phenol, and ethanol precipitated. Products of primer extension were analyzed on 8% sequencing polyacrylamide gels. A similar protocol was used for the sequencing of SP6 transcripts and splicing products generated during transient expression in HeLa cells except that Moloney murine leukemia virus reverse transcriptase was replaced with 20 U of avian myeloblastosis virus reverse transcriptase (Life Sciences) and appropriate dideoxynucleotide triphosphates were added at a concentration of 40 μM to each termination reaction.

RESULTS

In vitro splicing of pre-mRNAs with short internal exons. Plasmid pSP64HBΔ6 (hereafter referred to as HBΔ6), carrying the human β-globin gene under the control of the SP6 promoter (21), was modified by duplication of the globin segment spanning the first intron and adjacent sequences of both flanking exons. The resultant plasmid, DUP171, contained two identical introns and a 171-nucleotide-long central exon, composed of sequences from exon 1 and exon 2. Internal deletions originating at the unique *Nco*I site in the internal exon of clone DUP171 were used to generate a set of nested clones with the internal exons ranging in size from 104 to 23 nucleotides (Fig. 1). After cleavage at the *Bam*HI site, all plasmids were transcribed by SP6 RNA polymerase and the pre-mRNA transcripts were spliced in the nuclear extract from HeLa cells. It is important to note that as a consequence of the duplication step, all four splice sites, as well as their immediately adjacent sequences, were the same in all of the pre-mRNAs studied. Thus, the competition between splice sites was not biased by their preexisting context and was affected solely by the introduced sequence modifications.

Figure 2A presents a time course of in vitro splicing for pre-mRNA transcribed from the parental clone DUP171. The reaction follows a regular pathway, and the final product consisting of three spliced exons accumulates in high amounts after 100 min of incubation. The number of RNA bands generated during splicing of this pre-mRNA and the rate of their migration correlate well with the electrophoretic properties of the intermediates and final products predicted for the regular splicing pathway of a two-intronic pre-mRNA. Their structures are shown in Fig. 2B.

In vitro splicing of pre-mRNA transcribed from clone DUP51, containing a 51-nucleotide-long internal exon, generated the final product containing three exons and, in addition, approximately 20 to 25% of the product in which the internal exon had been skipped (Fig. 3). This is shown by the comigration of the latter product with the spliced product for monointronic HBΔ6 pre-mRNA (Fig. 3, lane 5). Skipping

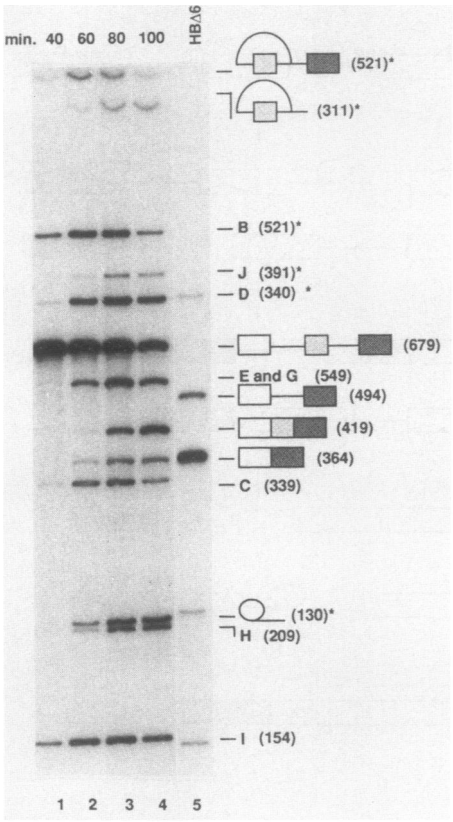

FIG. 3. Time course of the in vitro splicing reaction for DUP51 pre-mRNA. A 679-nucleotide substrate was spliced in vitro and analyzed as described for Fig. 2A. The products and intermediates are indicated on the right. The final product generated by skipping of the internal 51-nucleotide exon is indicated by its comigration with the final, 364-nucleotide-long product of in vitro splicing, obtained for HBΔ6 pre-mRNA (lane 5). The actual length of the skipping product is 368 nucleotides, as explained in the legend to Fig. 1.

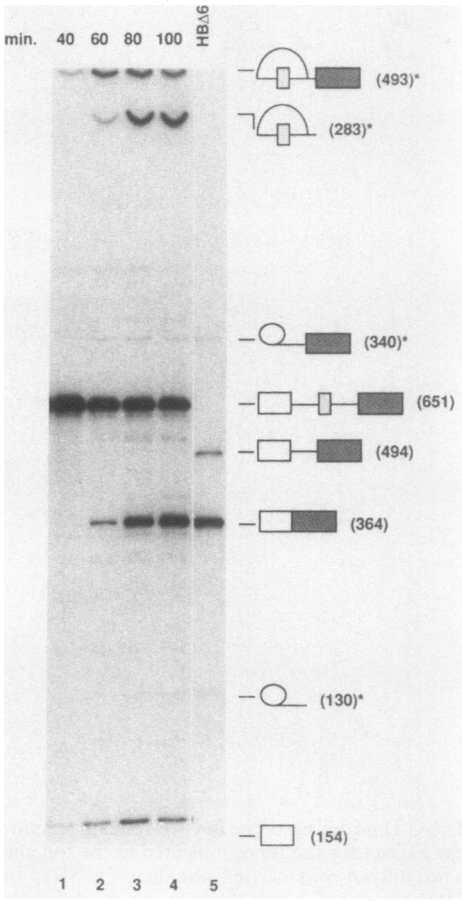

FIG. 4. Time course of the in vitro splicing reaction for DUP23 pre-mRNA. A 651-nucleotide-long substrate was spliced in vitro and analyzed as described for Fig. 2A. Structures and sizes (in nucleotides) of the intermediates and products generated by this pathway are depicted on the right. Other designations are as described for Fig. 2.

of the internal exon is also indicated by the appearance at the top of the gel of bands migrating at a much slower rate than the input pre-mRNA and other intermediates. Such a low electrophoretic mobility can be explained by the formation of large superlariats containing both introns and the skipped internal exon, as depicted on the right in Fig. 3. The structures of the superlariats were confirmed by primer extension and debranching analysis (42). The branch of the lariats is located within the downstream intron at the regular position 37 nucleotides upstream from the 3′ splice site (results not shown).

Skipping is the predominant reaction for pre-mRNA transcribed from clone DUP23, containing a 23-nucleotide-long internal exon (Fig. 4). The pattern of splicing for this transcript becomes very simple, lacking most of the intermediates and products seen in Fig. 2A, with the final spliced product comigrating with the product of splicing for monointronic HBΔ6 pre-mRNA (Fig. 4, lane 5). Skipping of the internal exon is also indicated, as in the case of DUP51 pre-mRNA, by the existence of slowly migrating large superlariat structures.

The results described above suggest that the length of the internal exon affects the course of the splicing reaction leading to alternative selection of splice sites. To investigate this effect in more detail, we extended the range of tested substrates to include transcripts containing internal exons 104, 87, 33, and 29 nucleotides long. Analysis of in vitro splicing of all pre-mRNAs, including those tested previously, is presented in Fig. 5. The transcripts containing internal exons 171, 104, and 87 nucleotides long are spliced by a pathway that generates only a fully spliced product containing all three exons (Fig. 5, lanes 1 to 3, respectively). As shown above (Fig. 3), splicing of DUP51 pre-mRNA generates two spliced products (Fig. 5, lane 4), whereas splicing of pre-mRNAs with shorter exons, i.e., 33, 29, and 23 nucleotides long, leads predominantly to exon skipping (Fig. 5, lanes 5 to 7, respectively). Note the existence of a band that migrates above (lanes 2 and 3) or below (lane 4) band C. Its unusual electrophoretic mobility, different for each clone tested, indicates a lariat intermediate, containing the first intron and the internal exon of variable length. This intermediate is generated when cleavage at both 5′ splice sites occurs simultaneously. This observation is consistent with earlier work (7).

Splicing of pre-mRNAs with short internal exons during transient expression in HeLa cells. Several reports showed that patterns of splicing in vitro and in vivo may differ significantly (1, 10, 44, 45) and that splice site selection in vitro may be affected by the concentration of the nuclear

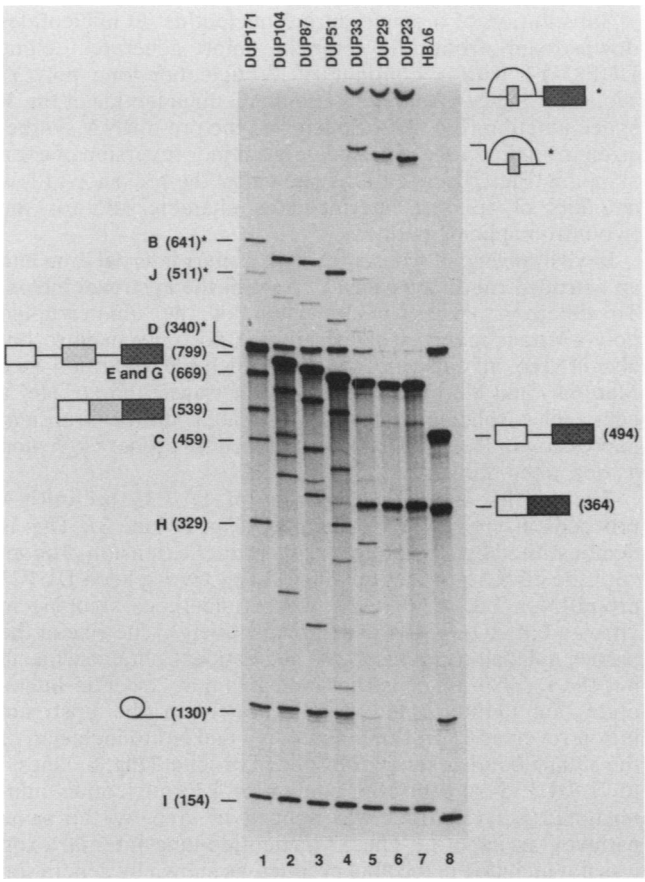

FIG. 5. In vitro splicing of pre-RNAs with internal exons 23 to 171 nucleotides long. All transcripts were spliced in vitro for 100 min and analyzed on an 8% polyacrylamide sequencing gel. Descriptions on the left indicate the splicing products and intermediates for DUP171 pre-mRNA; those on the right are for control, HBΔ6 pre-mRNA and for superlariats generated as a result of exon skipping. The name of each transcript above the lanes indicates the length of the internal exon in each pre-mRNA. Sizes are indicated in nucleotides.

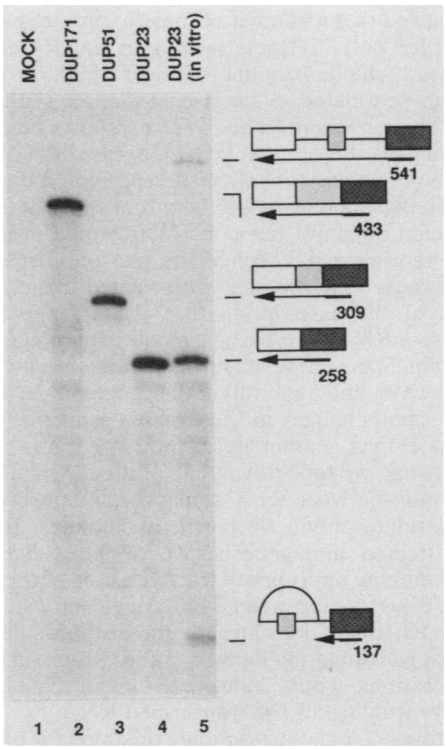

FIG. 6. Analysis of splicing pathways in vivo. Total RNA isolated from HeLa cells transfected with the DNA constructs indicated above the lanes was analyzed by primer extension as described in Materials and Methods. Structure of predicted templates for reverse transcription and sizes (in nucleotides) of the corresponding primer extension products are shown on the right. Lane 1 is a negative control carried out with RNA isolated from HeLa cells electroporated without exogenous DNA. As a control, primer extension products obtained on DUP23 pre-mRNA spliced in vitro are shown in lane 5.

splicing extract (19, 38). Thus, it was important to ascertain that the exon skipping described above is not an aberrant effect of in vitro splicing.

Three different clones, DUP171, DUP23, and DUP51, representing three modes of splicing (exon inclusion, exon skipping, and both pathways combined, respectively) were reconstructed (see Materials and Methods) and used for in vivo studies. These constructs, when transfected into HeLa cells, should generate pre-mRNAs with virtually the same size and nucleotide composition [except for the presence of poly(A) tails] as the SP6 transcripts used for in vitro studies, allowing for direct comparison of splicing pathways in vitro and in vivo.

Figure 6 presents results of primer extension analysis of in vivo-spliced RNA obtained after transient expression of clones DUP171, DUP51, and DUP23 in transfected HeLa cells. The electrophoretic mobilities of primer extension products obtained for clones DUP171 (Fig. 6, lane 2) and DUP51 (lane 3) suggested that endogenous precursors with internal exons 171 and 51 nucleotides long, respectively, are spliced in vivo via the regular pathway in which all three exons are included in the mature product. This observation was further confirmed by direct sequence analysis of the primer extension products (data not shown).

Analysis of the primer extension product for clone DUP23 revealed that, as in vitro, the 23-nucleotide-long internal exon was skipped in vivo. This skipping is indicated by the comigration of this extension product (Fig. 6, lane 4) with its equivalent reverse transcribed on the in vitro-spliced DUP23 pre-mRNA (lane 5) and by its nucleotide sequence (not shown). Even prolonged exposure of the autoradiogram did not show any indication of the regular splicing pathway proceeding by ligation of all three exons. We conclude that skipping of short internal exons takes place in vivo as well as in vitro, indicating that it is not an artifact of the in vitro splicing reaction.

In vitro splicing of transcripts with a short internal exon and a long uninterrupted poly(Y) tract in the upstream intron. The fact that short exons are skipped in vitro and in vivo was puzzling since a number of naturally occurring pre-mRNAs contain internal exons even shorter than those present in our model transcripts. For example, quail fast skeletal muscle troponin I pre-mRNA contains a 7-nucleotide-long exon (4), human renin pre-mRNA contains a 9-nucleotide-long exon (27), chicken pro-alpha 2(I) collagen pre-mRNA contains 11- and 18-nucleotide-long exons (2, 47), and rat fast skeletal troponin T pre-mRNA contains seven exons in the 12- to 18-nucleotide range (6). Inspection of sequences in these

genes revealed that a number of these exons are preceded by uninterrupted poly(Y) tracts ranging in length from 20 to 35 nucleotides, which are much longer than the 10 to 12 nucleotides postulated as a consensus length of this element (28, 43). This observation and recent reports on the role of the mammalian (11, 36) and yeast (34) poly(Y) sequences in the selection of acceptor splice sites prompted us to test the hypothesis that increasing the length of the poly(Y) tract in the upstream intron of our constructs would promote utilization of the internal 3′ splice site and prevent exon skipping. For these experiments, we selected clone DUP33 on the basis of the assumption that the splicing pattern of DUP33 pre-mRNA (exhibiting a small percentage of regular splicing; Fig. 5, lane 5) results from a subtle balance between both pathways and that this balance could be altered by relatively small changes in the existing sequence elements.

To create long uninterrupted poly(Y) tracts in the upstream introns, we took advantage of the fact that within the 37 nucleotides between the 3′ splice site and the branch point adenine residue, only 6 are purine nucleotides. By oligonucleotide-directed mutagenesis, we obtained five different clones containing appropriate transversions of these purines (Fig. 7A). Since we introduced only single base substitutions within the existing poly(Y) tracts, the distances between the splice sites remained unchanged, and it seems unlikely that these alterations would cause any major changes in the secondary structure of the transcribed RNA.

The pathways of in vitro splicing obtained for pre-mRNAs transcribed from these clones are shown in Fig. 7B. Substitution of all purines present in the original sequence, with the exception of one adenine located at position 30, resulted in an uninterrupted stretch of 28 pyrimidines directly adjacent to the branch point sequence. We obtained two variant clones with the same length and position of poly(Y) tract, DUP33-Y5 and DUP33-Y5′, differing only by the type of pyrimidine (C or T) used to replace the purines (Fig. 7A). Lanes 2 and 3 in Fig. 7B show the splicing pattern for pre-mRNAs transcribed from these two clones. In contrast to the pre-mRNA transcribed from the parental plasmid (Fig. 7B, lane 1), exon skipping is now almost completely reversed, as shown by the disappearance of the spliced product containing only exons 1 and 3 and by the lack of superlariat structures at the top of the gel. Furthermore, these changes are accompanied by the accumulation of RNA with the electrophoretic mobility predicted for the spliced product containing all three exons and by the appearance of the intermediates characteristic for a complex splicing pathway of two-intronic precursors. Although quantitative differences can be seen between the DUP33-Y5 and DUP33-Y5′ variants, both transcripts are now efficiently spliced via a regular pathway, regardless of the specific nature of pyrimidines used for substitution.

Substitution of 3 purine nucleotides at positions 17, 18, and 22 downstream from the branch point or 2 two purine nucleotides at positions 10 and 14 generated clones carrying a 15-nucleotide-long uninterrupted poly(Y) tract located either near the center of the sequence between the branch point and the 3′ splice junction (clone DUP33-Y3) or directly adjacent to the branch point sequence (clone DUP33-Y2; Fig. 7A). In vitro splicing of pre-mRNAs transcribed from these constructs exhibited a partial change in splice site selection (Fig. 7B, lanes 4 and 5). Both splicing pathways are now operating, generating spliced products with and without the internal exon at an approximately 1:1 ratio. In addition, there appears to be a decrease in the overall yield of the final spliced products.

Substitution of a single purine nucleotide 30 nucleotides downstream from the branch point generated clone DUP33-Y1 with a continuous 13-nucleotide-long poly(Y) sequence adjacent to the invariant AG dinucleotide at the 3′ splice junction (Fig. 7A). Splicing of the pre-mRNA synthesized from this clone revealed only a minor reversion of exon skipping (Fig. 7B, lane 6), as shown by the presence of low amounts of splicing intermediates characteristic for the two-intron splicing pathway.

In vivo splicing of transcripts with a short internal exon and an extended continuous poly(Y) tract in the upstream intron. To determine whether elongation of the uninterrupted poly(Y) tract reverses exon skipping not only in vitro but also in vivo, all constructs with modified poly(Y) tracts (see Materials and Methods) were used for transfection of HeLa cells. The splicing products generated during transient expression in HeLa cells were analyzed by primer extension as described above.

The results show that splicing of DUP33 pre-mRNA proceeds in vivo by exon skipping (Fig. 8, lane 3). This is demonstrated by comigration of its primer extension product with the cDNA product obtained for in vivo-spliced DUP23 pre-mRNA (Fig. 8, lane 1), which undergoes skipping in vitro and in vivo (Fig. 4 and 6, respectively). The size of the primer extension product (291 nucleotides) obtained for all but the DUP33-Y1 construct indicates that 28- or 15-nucleotide-long uninterrupted poly(Y) tracts in the upstream intron reversed exon skipping in vivo and led to inclusion of the 33-nucleotide exon in the spliced product (Fig. 8, lanes 4 to 7). DUP33-Y1 transcript containing a 13-nucleotide uninterrupted poly(Y) tract was spliced in vivo by the same pathway as in vitro. The 33-nucleotide-long internal exon was not included in the final product, as shown by generation of a 258-nucleotide cDNA (Fig. 8, lane 8).

DISCUSSION

Comparison of splicing pathways for several pre-mRNAs containing two identical introns and an internal exon ranging in size from 23 to 171 nucleotides revealed that there is a strong relationship between the length of this exon and the utilization of the internal 3′ and 5′ splice sites. Skipping of the internal exon was observed during in vitro splicing of pre-mRNAs with internal exons 23, 29, and 33 nucleotides long and to a limited extent for a pre-mRNA with a 51-nucleotide exon. In contrast, exons containing 87, 104, and 171 nucleotides were included in the spliced product without generation of detectable amounts of a skipping product. Whenever exon skipping was observed, splicing proceeded by formation of a superlariat structure with a branch point at the regular position, 37 nucleotides upstream from the 3′ splice site, located within the downstream intron. Note that because of the different endpoints of the internal deletions (data not shown), the three pre-mRNAs which exhibited exon skipping as a major splicing pathway differed partially in the sequence of the central exon. Thus, in our model transcripts, exon skipping does not seem to be promoted by a change in a specific sequence within the internal exon.

Skipping of short internal exons is not an artifact of in vitro splicing. The transient expression assays revealed that internal exons 23 and 33 nucleotides long are skipped also in vivo. It was previously proposed that progressive exposure of endogenously synthesized RNAs to the nuclear environment favors correct splice site selection, i.e., pairing of a 5′ splice site with the next available 3′ splice site (1). It appears, however, that the principle of "first come-first

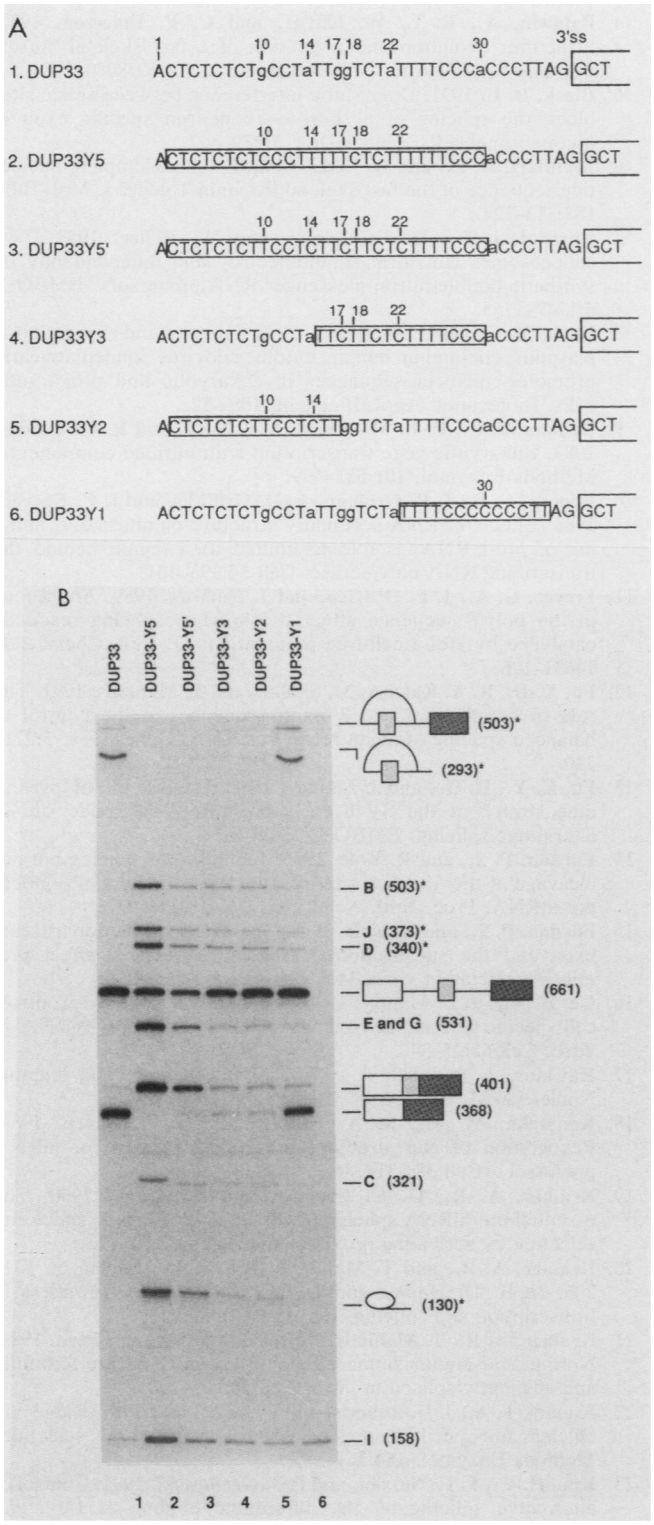

FIG. 7. (A) Nucleotide sequence of the poly(Y) tract in the parental DUP33 clone and in variant clones generated by site-directed mutagenesis. All purines, located in the original poly(Y) tract at positions 10, 14, 17, 18, 22, and 30 relative to the branch point adenine, are shown in lowercase letters. The beginning of the second exon and the position of the 3′ splice site (3′ss) are shown. Substitution of individual purines with cytosine or thymine nucleotides leads to formation of uninterrupted pyrimidine stretches, indicated by close boxes. (B) In vitro splicing reaction with pre-mRNAs containing modified poly(Y) tracts. The pre-mRNAs tested are indicated above the lanes. Splicing was carried out in HeLa cell nuclear extract for 100 min, and the RNA was isolated and analyzed on an 8% polyacrylamide sequencing gel. Structure of the products and intermediates are either depicted at the right or illustrated in Fig. 2B. Splicing of the parental pre-mRNA, DUP33 is shown in lane 1. Sizes are indicated in nucleotides.

serve" does not apply to our model DUP23 and DUP33 pre-mRNAs.

Why are the short internal exons skipped both in vitro and in vivo although they are preceded by sequence elements identical to those in the downstream intron? It seems likely that the failure to recognize short internal exons is due to the juxtaposition of 3′ and 5′ splice sites flanking the internal exon and consequent steric hindrance in the binding of the splicing factors to these sites. The idea that short internal exons cannot be efficiently spliced is also supported by statistical data (17, 29, 48). In vertebrates, approximately 85% of the 1,300 internal exons surveyed fall into the range of 50 to 200 nucleotides in length. Internal exons shorter than 50 nucleotides are rare, constituting no more than 4% of all internal exons (17).

Our results suggest that the improvement in the strength of the internal 3′ splice site provided by the long uninterrupted poly(Y) tracts can minimize the steric hindrance effect and prevent exon skipping. This was clearly seen in vitro and in vivo for clones with an uninterrupted poly(Y) tract containing 28 contiguous pyrimidines adjacent to the branch point. The 15-nucleotide-long poly(Y) tract, whether immediately adjacent to the branch point or located further downstream, led to partial reversal of exon skipping in vitro and to complete reversal during in vivo splicing. Exon skipping was not reversed either in vitro or in vivo by the 13-nucleotide-long poly(Y) tract adjacent to the 3′ splice site. Although skipping reversal apparently depends strongly on the length of the uninterrupted poly(Y) tract, a positional component may also play some role since the shortest tract was also the one closest to the 3′ splice site. The latter result is consistent with the observation that purines inserted close to the invariant AG dinucleotide in the 3′ splice site had a smaller effect than those close to the branch point on the utilization of the splice site (36). Skipping of short exons seems also to be reversible by the increase in the concentration of the splicing factors. This is seen by a shift toward exon inclusion during splicing of DUP51, DUP33-Y3, and DUP33-Y2 pre-mRNAs in vivo, presumably as a result of high local concentrations of the splicing factors in intact nuclei. For DUP51 pre-mRNA, we have also observed that the balance between fully spliced or skipped products can be affected in vitro by the concentration of the extracts (not shown). Effects of the concentration of the extracts or purified splicing factors on splice site selection have been reported elsewhere (16, 19, 38).

Are the long uninterrupted poly(Y) tracts the only sequence elements that prevent skipping of short internal exons? Although, as mentioned earlier, a number of short exons are preceded by long uninterrupted poly(Y) tracts, there are several short internal exons that do not share this characteristic (2, 6, 47). Other sequence elements may play a role in these pre-mRNAs, as suggested by the recent observations that mutations in the internal 5′ splice site affect splicing of the adjacent upstream exon in various pre-mRNAs (23, 40, 46).

Nevertheless, it should be stressed that the substitution of

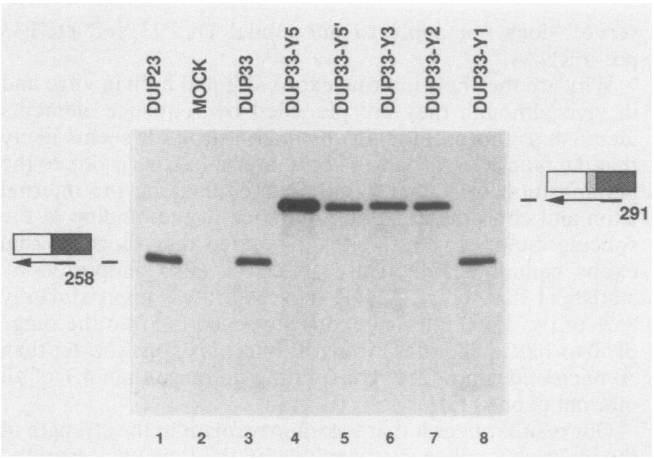

FIG. 8. Analysis of in vivo splicing pathways for pre-mRNAs containing a modified poly(Y) tract. Total RNA isolated from HeLa cells transfected with the DNA constructs indicated above the lanes was analyzed by primer extension as described in Materials and Methods. The predicted structures of templates for reverse transcription and the lengths (in nucleotides) of the corresponding primer extension products are shown next to the autoradiogram. The mock control (lane 2) was obtained by primer extension performed with the RNA isolated from HeLa cells electroporated without exogeneous plasmid DNA.

only two purines in DUP33-Y2 pre-mRNA, which generated an uninterrupted poly(Y) tract of moderate length, had a dramatic effect on splice site selection in vivo, leading to complete reversal of exon skipping. This and other results presented in this report suggest that individual purine nucleotides located in the poly(Y) tracts may serve as an important element modulating the strength of the 3′ splice sites and in conjunction with the length of the immediately adjacent exons may determine patterns of splice site selection in multi-intron precursors.

After this work was completed, Black (5) reported that the 18-nucleotide-long internal exon from a mouse c-*src* gene is included in neural cells but skipped in cells of other than neural origin. Increasing the length of the exon to 109 nucleotides led to its efficient inclusion in HeLa cells, which fail to recognize the 18-nucleotide exon, indicating the role of exon length in splice site selection.

ACKNOWLEDGMENTS

We thank Elizabeth Smith for excellent technical assistance, Chris Joneckis for construction, and Derk Schultz for preliminary characterization of the DUP171 clone. We are very grateful to Eng-Shang Huang from the Lineberger Comprehensive Cancer Center for the DNA clone carrying the human cytomegalovirus immediate-early promoter. We also thank our colleagues for comments on the manuscript.

This work was supported in part by grants from the National Institutes of Health and American Cancer Society to R.K.

REFERENCES

1. **Aebi, M., H. Horning, R. A. Padgett, J. Reiser, and C. Weissmann.** 1986. Sequence requirement for splicing of higher eucaryotic nuclear pre-mRNA. Cell **47**:555–565.
2. **Aho, S., V. Tate, and H. Boedtker.** 1984. Location of the 11 bp exon in the chicken pro-alpha2(I) collagen gene. Nucleic Acids Res. **15**:6117–6125.
3. **Andreadis, A., M. E. Gallego, and B. Nadal-Ginard.** 1987. Generation of protein isoform diversity by alternative splicing. Annu. Rev. Cell Biol. **3**:207–242.
4. **Baldwin, A., E. L. W. Kittler, and C. P. Emerson.** 1985. Structure, evolution and regulation of a fast skeletal muscle troponin I gene. Proc. Natl. Acad. Sci. USA **82**:8080–8084.
5. **Black, D. L.** 1991. Does steric interference between splice sites block the splicing of a short c-src neuron specific exon in non-neuronal cells? Genes Dev. **5**:389–402.
6. **Breitbart, R. E., and B. Nadal-Ginard.** 1986. Complete nucleotide sequence of the fast skeletal troponin T gene. J. Mol. Biol. **188**:313–324.
7. **Christofori, G., D. Frendewey, and W. Keller.** 1987. Two spliceosomes can form simultaneously and independently on synthetic double-intron messenger RNA precursors. EMBO J. **6**:1747–1755.
8. **Davis, M. G., and E.-S. Huang.** 1988. Transfer and expression of plasmids containing human cytomegalovirus immediate-early promoter-enhancer sequences in eukaryotic and prokaryotic cells. Biotechnol. Appl. Biochem. **10**:6–12.
9. **Dignam, J. D., P. L. Martin, B. S. Shastry, and R. G. Roeder.** 1983. Eukaryotic gene transcription with purified components. Methods Enzymol. **101**:582–599.
10. **Eperon, L. P., I. R. Graham, A. D. Griffiths, and I. C. Eperon.** 1988. Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase. Cell **54**:393–401.
11. **Freyer, G. A., J. P. O'Brien, and J. Hurwitz.** 1989. Alterations in the polyY sequence affect the in vitro splicing reactions catalyzed by HeLa cell-free preparations. J. Biol. Chem. **264**: 14631–14637.
12. **Fu, X.-D., R. A. Katz, A. M. Skalka, and T. Maniatis.** 1991. The role of branchpoint and 3′-exon sequences in the control of balanced splicing of avian retrovirus RNA. Genes Dev. **5**:211–220.
13. **Fu, X.-Y., H. Ge, and J. Manley.** 1988. The role of polypyrimidine stretch at the SV40 early pre-mRNA 3′ splice site in alternative splicing. EMBO J. **7**:809–817.
14. **Furdon, P. J., and R. Kole.** 1986. Inhibition of splicing but not cleavage at the 5′ splice site by truncating the human β-globin pre-mRNA. Proc. Natl. Acad. Sci. USA **83**:27–931.
15. **Furdon, P. J., and R. Kole.** 1988. The length of the downstream exon and the substitution of specific sequences affect pre-mRNA splicing in vitro. Mol. Cell. Biol. **8**:860–866.
16. **Ge, H., and J. L. Manley.** 1990. A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. Cell **62**:25–34.
17. **Hawkins, J. D.** 1988. A survey of intron and exon lengths. Nucleic Acids Res. **16**:9893–9908.
18. **Konarska, M. M., R. A. Padgett, and P. A. Sharp.** 1984. Recognition of cap structure in splicing in vitro of mRNA precursors. Cell **38**:731–736.
19. **Krainer, A. R., G. C. Conway, and D. Kozak.** 1990. The essential pre-mRNA splicing factor SF2 influences 5′ splice site selection by activating proximal sites. Cell **62**:35–42.
20. **Krainer, A. R., and T. Maniatis.** 1988. RNA splicing, p. 131–206. *In* B. D. Hames and D. M. Glover (ed.), Frontiers in transcription and splicing. IRL Press, Oxford.
21. **Krainer, A. R., T. Maniatis, B. Ruskin, and M. R. Green.** 1984. Normal and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced in vitro. Cell **36**:993–1005.
22. **Kunkel, T. A., J. D. Roberts, and R. A. Zakour.** 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. **154**:367–382.
23. **Kuo, H.-C., F. H. Nassim, and P. J. Grabowski.** 1991. Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. Science **251**:1045–1050.
24. **Lear, A. L., L. P. Eperon, I. M. Wheatley, and I. C. Eperon.** 1990. Hierarchy of 5′ splice site preference determined in vivo. J. Mol. Biol. **211**:103–115.
25. **Maniatis, T.** 1991. Mechanism of alternative pre-mRNA splicing. Science **251**:33–34.
26. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 188–209. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Miyazaki, H., et al.** 1984. Structure of the human renin gene.

Proc. Natl. Acad. Sci. USA **81**:5999–6003.
28. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–467.
29. **Naora, H., and N. J. Deacon.** 1982. Relationship between the total size of exons and introns in protein-coding genes of higher eukaryotes. Proc. Natl. Acad. Sci. USA **79**:6196–6200.
30. **Nelson, K. K., and M. R. Green.** 1990. Mechanism for cryptic splice site activation during pre-mRNA splicing. Proc. Natl. Acad. Sci. USA **87**:6253–6257.
31. **Ohshima, Y., and Y. Gotoh.** 1987. Signals for the selection of a splice site in pre-mRNA. Computer analysis of splice junction sequences and like sequences. J. Mol. Biol. **195**:247–259.
32. **Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler, and P. A. Sharp.** 1986. Splicing of messenger RNA precursors. Annu. Rev. Biochem. **55**:1119–1150.
33. **Parent, A., S. Zeitlin, and A. Efstratiadis.** 1987. Minimal exon sequence splicing requirements for efficient in vitro splicing of monointronic nuclear pre-mRNA. J. Biol. Chem. **262**:11284–11291.
34. **Patterson, B., and C. Guthrie.** 1991. A U-rich tract enhances usage of an alternative 3′ splice site in yeast. Cell **64**:181–187.
35. **Potter, H.** 1987. Introduction of DNA into mammalian cells, p. 9.3.1–9.3.3. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology. J. Wiley & Sons, New York.
36. **Reed, R.** 1989. The organization of 3′ splice-site sequences in mammalian introns. Genes Dev. **3**:2113–2123.
37. **Reed, R., and T. Maniatis.** 1985. Intron sequences involved in lariat formation during pre-mRNA splicing. Cell **41**:95–105.
38. **Reed, R., and T. Maniatis.** 1986. A role for exon sequences and splice-site proximity in splice site selection. Cell **46**:681–690.
39. **Reed, R., and T. Maniatis.** 1988. The role of the mammalian branchpoint sequence in pre-mRNA splicing. Genes Dev. **2**:1268–1276.
40. **Robberson, B. L., G. J. Cote, and S. M. Berget.** 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. **10**:84–94.
41. **Ruskin, B., and M. R. Green.** 1985. Role of the 3′ splice site consensus sequence in mammalian pre-mRNA splicing. Nature (London) **317**:732–734.
42. **Ruskin, B., and M. R. Green.** 1985. An RNA processing activity that debranches RNA lariats. Science **229**:135–140.
43. **Sharp, P. A.** 1987. Splicing of messenger RNA precursors. Science **235**:766–771.
44. **Solnick, D.** 1985. Alternative splicing caused by RNA secondary structure. Cell **43**:667–676.
45. **Solnick, D., and S. I. Lee.** 1987. Amount of RNA secondary structure required to induce an alternative splice. Mol. Cell. Biol. **7**:3194–3198.
46. **Talerico, M., and S. M. Berget.** 1990. Effect of 5′ splice site mutations on splicing of the preceding intron. Mol. Cell. Biol. **10**:6299–6305.
47. **Tate, V. E., M. H. Finer, H. Boedtker, and P. Doty.** 1983. Chick pro-alpha2(I) collagen gene: exon location and coding potential for prepropeptide. Nucleic Acids Res. **11**:91–104.
48. **Traut, T. W.** 1988. Do exons code for structural or functional units in proteins? Proc. Natl. Acad. Sci. USA **85**:2944–2948.
49. **Treisman, R., S. H. Orkin, and T. Maniatis.** 1983. Specific transcription and RNA splicing defects in five cloned β-thalassemia genes. Nature (London) **302**:591–596.
50. **Treisman, R., S. H. Orkin, and T. Maniatis.** 1983. Structural and functional defects in β-thalassemia, p. 99–121. *In* G. Stamatoyannopoulos and A. W. Nienhuis (ed.), Globin gene expression and hematopoietic differentiation. Alan R. Liss, New York.
51. **Treisman, R., N. J. Proudfoot, M. Shander, and T. Maniatis.** 1982. A single-base change at a splice site in a β^0-thalassemic gene causes abnormal RNA splicing. Cell **29**:903–911.
52. **Turnbull-Ross, A. D., A. J. Else, and I. C. Eperon.** 1988. The dependence of splicing efficiency on the length of 3′ exon. Nucleic Acids Res. **16**:395–411.
53. **Wieringa, B., F. Meyer, J. Reiser, and C. Weissmann.** 1983. Unusual splice sites revealed by mutagenic inactivation of an authentic splice site of the rabbit β-globin gene. Nature (London) **301**:38–43.
54. **Zhuang, Y., A. M. Goldstein, and A. M. Weiner.** 1989. UAC UAAC is the preferred branch site for mammalian mRNA splicing. Proc. Natl. Acad. Sci. USA **86**:2752–2756.
55. **Zhuang, Y., H. Lueng, and A. M. Weiner.** 1987. The natural 5′ splice site of simian virus 40 large T antigen can be improved by increasing the base complementarity to U1 RNA. Mol. Cell. Biol. **7**:3018–3020.
56. **Zhuang, Y., and A. M. Weiner.** 1986. A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation. Cell **46**:827–835.
57. **Zhuang, Y., and A. M. Weiner.** 1989. A compensatory base change in human U2 snRNA can suppress a branch site mutation. Genes Dev. **3**:1545–1552.